Critical Evidence 2.2.2

The Redesigned SAT® Pilot Predictive Validity Study: A First Look

RESEARCH REPORT 2016-1

CollegeBoard

The Redesigned SAT® Pilot Predictive Validity Study: A First Look

By Emily J. Shaw, Jessica P. Marini, Jonathan Beard, Doron Shmueli, Linda Young, and Helen Ng

Emily J. Shaw is a senior director in the College Board's research department.

Jessica P. Marini is an assistant research scientist in the College Board's research department.

Jonathan Beard is an associate research scientist in the College Board's research department.

Doron Shmueli is a research statistician II in the College Board's research department.

Linda Young is a senior database statistician in the College Board's research department.

Helen Ng is a senior research analyst in the College Board's research department.

About the College Board

The College Board is a mission-driven not-for-profit organization that connects students to college success and opportunity. Founded in 1900, the College Board was created to expand access to higher education. Today, the membership association is made up of over 6,000 of the world's leading educational institutions and is dedicated to promoting excellence and equity in education. Each year, the College Board helps more than seven million students prepare for a successful transition to college through programs and services in college readiness and college success — including the SAT® and the Advanced Placement Program®. The organization also serves the education community through research and advocacy on behalf of students, educators, and schools. For further information, visit www.collegeboard.org.

VALIDITY

For more information on College Board research and data, visit research.collegeboard.org.

Contents

Tables

Figures

Executive Summary

The College Board conducted a pilot predictive validity study to provide colleges and universities with early information about the relationship between the redesigned SAT® and college grades. Fifteen four-year institutions were recruited to administer a pilot form of the redesigned SAT to between 75 and 250 first-year, first-time students very early in the fall semester of 2014. Measures were taken to ensure that the redesigned SAT was administered to students under standardized conditions and that students were motivated to perform well on the test. In June 2015, participating institutions provided the College Board with first-year performance data for those students participating in the fall 2014 administration of the redesigned SAT so that relationships between SAT scores and college performance could be analyzed. Results of study analyses show that the redesigned SAT is as predictive of college success as the current SAT, that redesigned SAT scores improve the ability to predict college performance above high school GPA alone, and that there is a strong, positive relationship between redesigned SAT scores and grades in matching college course domains, suggesting that the redesigned SAT is sensitive to instruction in English language arts, math, science, and history/social studies.

Introduction

In February of 2013, the College Board announced it would undertake a redesign of the SAT® in order to develop an assessment that better reflects the work that students will do in college, focusing on the core knowledge and skills that evidence has shown to be critical in preparation for college and career. The redesigned test will be introduced in March 2016 and will include a number of important changes (for a full description of these changes, visit https://collegereadiness.collegeboard.org/sat/test-design).

The redesigned SAT Evidence-Based Reading and Writing section and (optional) Essay portions will incorporate key design elements supported by evidence, including:

- The use of a range of text complexity aligned to college- and career-ready reading levels;
- An emphasis on the use of evidence and source analysis;
- The incorporation of data and informational graphics that students will analyze along with text;
- A focus on relevant words in context and on word choice for rhetorical effect;
- Attention to a core set of important English language conventions and to effective written expression; and
- The requirement that students interact with texts across a broad range of disciplines.

The key evidence-based design elements that will be incorporated into the redesigned SAT Math section include:

- A focus on the content that matters most for college and career readiness (rather than a vast array of concepts);
- An emphasis on problem solving and data analysis; and
- The inclusion of "Calculator: Permitted" questions as well as "Calculator: Not Permitted" questions and attention to the use of the calculator as a tool.

Instead of the SAT having three sections, each on a 200–800 scale, the redesigned SAT will now have two broad sections: Evidence-Based Reading and Writing, and Math, each on a 200–800 scale. Within the Evidence-Based Reading and Writing section, there will be two test scores: a Reading Test score and a Writing and Language Test score, each on a 10–40 scale. The Math section will also produce a Math Test score on a 10–40 scale. The Essay will now be optional, and students will have 50 minutes instead of 25 minutes to write. There will also be a number of subscores and cross-test scores produced to provide richer information to students, schools, and institutions on student performance. Another notable change is that students will earn points for the questions they answer correctly and will not lose points for incorrect answers as they had on the previous SAT.

As with the redesign of all assessments, it is important to examine and understand how the changes to the content and format of the test impact the inferences made from the test's scores for their intended uses. One primary use of the SAT is for admission and placement

decisions and, therefore, it was important to examine the relationship between the scores from the redesigned test with college outcomes such as first-year grade point average (FYGPA) and college course grades. In order to conduct such an analysis a pilot study was initiated because the test is not yet operational.

This paper describes the research efforts and the results of the first predictive validity study on a pilot form of the redesigned SAT. The findings should inform the higher education community with regard to any expected changes in the predictive validity of the redesigned SAT in college admission.

Methodology

Study Design

A typical operational admission validity study would use students' recorded SAT scores and their reported FYGPA to examine the statistical association between the two. Because this was a pilot study and not an operational validity study, it was necessary to first administer a pilot form of the redesigned SAT to students who had just begun their first year of college. We would then follow those students through their first year of college and collect their grades and FYGPA as the outcome for analyses. In order to do this, the College Board partnered with four-year institutions in the U.S. to administer the test and then collect student grades. The general process for institutional participation was to:

- Determine a preferred date and time early in the first semester to hold the test administration.
- Recruit between 75 and 250 students to participate in the study. For students to be eligible to sit for the test/participate in the study, they had to be first-time, first-year students who had previously taken the SAT.
    - Students received a $100 gift card for participating in the study immediately following their test participation. To increase test-taking motivation, students were also made aware that they would receive a $50 gift card, mailed at a later date, if their scores on the redesigned SAT met or exceeded their most recent SAT scores on record at the College Board.
- Reserve a testing room(s) based on planned recruitment numbers and SAT room requirements/sample seating plans.
- Assist with the handling of test materials and test day administration (along with College Board and ETS staff).
- Deliver the student data file to the College Board at the end of the 2014–15 school year with student participants' first-year course work and grades.

Institutions were asked to assign a Study Coordinator as the point person for the study. The Study Coordinator or the institution was eligible to receive a fee for coordinating the study. In addition, each participating institution received an institution-specific validity report based on its data.

Participants

Institutional Sample

The goal for this study was to recruit 10 to 15 diverse four-year institutions for participation so that students could then be recruited to participate in a campus administration of the redesigned SAT. To design a sampling plan, we first outlined the population of four-year institutions from the College Board's *Annual Survey of Colleges* (ASC) from 2012, which collects information from colleges, universities, vocational/technical, and graduate schools that is of interest to potential applicants. The population of four-year institutions from which we would sample was specified as follows:

1. Located within the United States;
2. Accredited by at least one accrediting agency;
3. Has at least 200 enrolled degree-seeking, first-year students who sent SAT scores to the institution;
4. Uses test information to make admission decisions;
5. Is either public or private (but not private, for-profit); and
6. Is a bachelor's-degree-granting institution.

Based on these criteria, the number of total eligible institutions from which to sample was 699. Institutions were then stratified by region, admission selectivity, institution size, and institution control (public or private) to determine sample targets. The desired sample of institutions was then developed to best reflect the population while also aiding in the study administration (e.g., larger institutions would have a more likely chance of recruiting students to participate in the study). The recruitment of institutions was facilitated by regional College Board staff who are closely connected to colleges and universities. As the requirements for study participation were too burdensome for some institutions, similar institutions were identified as backup institutions in order to maintain as diverse and representative a sample of institutions as possible when selecting 10 to 15 institutions out of 699.

Table 1 provides information that can be used to compare the sample of study institutions to the population of institutions for recruitment. In addition, a comparison to the institutions included in the most recent national SAT Validity Study (Beard & Marini, 2015) is also provided. The College Board routinely conducts national validity research on the SAT, and sample comparisons to the institutional sample in these validity studies could aid in our understanding of comparisons of results in the current study to earlier validity research results. These sample comparisons show that the Pilot Study Sample includes more Southern and Southwestern institutions than the population and 2012 Validity Study Sample, fewer Mid-Atlantic institutions than the population or 2012 Validity Study Sample, and fewer Midwestern institutions than the 2012 Validity Study Sample. The Pilot Study Sample included more public institutions than private institutions (67% versus 33%), and this represents more public institutions and fewer private institutions than are in the population or the 2012 Validity Study Sample. With regard to selectivity, while the institutions that admit over 75% of applicants were well represented in the Pilot Study Sample (20%) as compared to the population (22%) and the 2012 Validity Study Sample (21%), the Pilot Study Sample included fewer institutions that admitted

between 50% and 75% of applicants (40%), and more institutions that admitted under 50% of applicants (40%) than the population or 2012 Validity Study Sample.

Table 1.

Comparison of Institutional Study Sample to Population of Institutions for Recruitment and Previous SAT Validity Research Sample

		Pilot Study Sample (n_i = 15)		Population (n_i = 699)		2012 SAT Validity Study Sample (n_i = 165)	
		n_i	%	n_i	%	n_i	%
U.S. Region	Midwest	1	7	73	10	26	16
	Mid-Atlantic	2	13	228	33	45	27
	New England	2	13	91	13	17	10
	South	5	33	150	21	31	19
	Southwest	3	20	48	7	20	12
	West	2	13	109	16	26	16
Control	Public	10	67	305	44	78	47
	Private	5	33	394	56	87	53
Admittance Rate	Under 50%	6	40	179	26	36	22
	50% to 75%	6	40	361	52	92	56
	Over 75%	3	20	152	22	35	21
Undergraduate Enrollment	Small	0	0	150	21	33	20
	Medium	5	33	332	48	66	40
	Large	2	13	106	15	33	20
	Very Large	8	53	110	16	33	20

Note. Percentages may not sum to 100 due to rounding. The population was based on four-year institutions from the College Board's *Annual Survey of Colleges* (ASC) from 2012, and criteria for inclusion were: located within the United States; accredited by at least one accrediting agency; has at least 200 enrolled degree-seeking, first-year students who sent scores to the institution; uses test information to make admission decisions; is either public or private (but not private, for-profit); and is a bachelor's-degree-granting institution. Undergraduate enrollment was categorized as follows: small – 750 to 1,999; medium – 2,000 to 7,499; large – 7,500 to 14,999; and very large – 15,000 or more.

Student Sample

Participating institutions were charged with recruiting as representative a sample of first-year students at their institution as possible. Students also had to have previously taken the SAT (in the 2014 College-Bound Seniors cohort) so that comparisons between their operational SAT scores and their redesigned SAT scores could be made. This comparison was primarily used to identify students with particularly low motivation during the pilot test administration.

There were 2,182 students who participated in the redesigned SAT test administration across the 15 institutions in the study. There were 32 students who were dropped from the sample

because they: (1) did not have an SAT score in the 2014 College-Bound Seniors cohort (n = 21); (2) did not have an SAT score on record at all (n = 5); (3) were determined not to be a first-time freshman (n = 1); or (4) could not be matched from their test administration data to the College Board database (n = 5).

Among the 2,150 students who remained in the sample, there was additional filtering that needed to take place to ensure that all students had the study variables of interest. There were 61 students who either did not have a high school grade point average (HSGPA) on record (n = 57) or who did not have a FYGPA (n = 4), and these students were removed from the study sample.

For the 2,089 remaining students, it was important to examine concerns with low motivation as this test was administered as part of a study as opposed to an actual SAT test associated with high stakes. First, operational SAT scores on record for students in the sample were concorded to redesigned SAT scores using a concordance table linking scores on both tests. The concordance table was developed by the Psychometrics team at the College Board. The difference between the actual redesigned SAT score that the student received for the study and the concorded redesigned SAT score was calculated, and this difference value was then standardized for each student (the student's difference score minus the mean difference score, divided by the standard deviation of the difference score). This was done for the Evidence-Based Reading and Writing (EBRW) section and the Math section. Standardized score differences that were greater than ±2 in either section were flagged. Another flag for low effort was created for students responding with "Disagree" or "Strongly Disagree" to the following statement on the redesigned SAT answer sheet, "I plan to put forth my best effort during the test today." The researchers determined that those students with an EBRW score difference flag and a Math score difference flag should be dropped from the study. Also, those with either an EBRW score difference flag (but no Math score difference flag) and a low effort flag were dropped, as well as students with a Math score difference flag (but no EBRW score difference flag) and a low effort flag. There were 39 students removed from the study based on these analyses. Therefore, the final sample included 2,050 students.

See Table 2 for a comparison of the student study sample to the population of students in the College-Bound Seniors 2014 cohort (College Board, 2014). In addition, a comparison to the students included in the most recent national SAT Validity Study is provided so that similarities and differences could be noted between the sample that is typically studied in the College Board's national, operational validity research to the smaller sample in this pilot validity study. These sample comparisons show that the Pilot Study Sample included more female students (64%) than either the population (53%) or the 2012 Validity Study Sample (55%), and also included more Asian students (20%) than the population (12%) or the 2012 Validity Study Sample (11%). The Pilot Study Sample essentially matched the population with regard to African American (13% for both), Hispanic (17% and 18%, respectively), and white (46% and 49%, respectively) students. The 2012 Validity Study Sample tended to include more white students and fewer underrepresented minority students than the Pilot Study Sample or the population.

Table 2.

Comparison of Student Sample to Population of Students and Previous SAT Validity Research Student Sample

		Pilot Study Sample (n_s = 2,050)		2014 College-Bound Seniors (n_s = 1,672,395)		2012 SAT Validity Study Sample (n_s = 223,109)	
		n	%	*n*	%	*n*	%
Gender	Male	743	36	783,570	47	100,739	45
	Female	1,306	64	888,825	53	122,370	55
Race/ Ethnicity	African American	262	13	212,524	13	19,326	9
	American Indian	9	0	9,767	1	991	0
	Asian	403	20	206,564	12	25,399	11
	Hispanic	358	17	300,357	18	24,787	11
	Other	64	3	64,774	4	6,135	3
	White	949	46	822,821	49	144,464	65
	Not Stated	5	0	55,588	3	2,007	1
Best Language	English Only	1,653	81	1,246,019	75	190,113	85
	English and Another	361	18	312,316	19	28,411	13
	Another Language	26	1	66,082	4	3,856	2
	Not Stated	10	0	47,978	3	729	0
Parental Income	< $40,000	284	14	279,901	17	19,820	9
	$40,000–$80,000	309	15	255,523	15	25,308	11
	$80,000–$120,000	286	14	203,870	12	24,714	11
	$120,000–$160,000	126	6	92,848	6	12,199	5
	$160,000–$200,000	74	4	49,211	3	6,696	3
	> $200,000	102	5	74,838	4	12,516	6
	Not Stated	869	42	716,204	43	121,856	55
Highest Parental Education Level	No High School Diploma	86	4	100,705	6	7,314	3
	High School Diploma	420	20	440,908	26	44,289	20
	Associate Degree	121	6	125,781	8	14,802	7
	Bachelor's Degree	697	34	484,624	29	78,556	35
	Graduate Degree	692	34	377,443	23	65,745	29
	Not Stated	34	2	142,934	9	12,403	6

Note. Percentages may not sum to 100 due to rounding. One student in the pilot study sample did not indicate gender.

Measures

Redesigned SAT Scores. Redesigned SAT scores were obtained from the special administrations of a pilot form of the redesigned SAT in the fall of 2014 for this study. This includes the following scores:

Two section scores (200 to 800 scale)

Evidence-Based Reading and Writing (not including SAT Essay) – increments of 10

Math – increments of 10

Three test scores (10 to 40 scale)

Reading – increments of 1

Writing and Language (not including SAT Essay) – increments of 1

Math – increments of 0.5

Two cross-test scores (10 to 40 scale)

Analysis in Science – increments of 1

Analysis in History/Social Studies – increments of 1

Redesigned SAT Pilot Study Questionnaire Responses. Self-reported responses to questions on test day informed this study design, including questions related to motivation and effort, as well as student information allowing researchers to match data from the pilot study to students' operational SAT scores on record.

SAT Questionnaire Responses. Self-reported gender, race/ethnicity, best language, parental education level, and parental income level were obtained from the SAT Questionnaire that students complete during registration for the operational SAT.

High School GPA. Self-reported HSGPA was obtained from the SAT Questionnaire when students had taken the operational SAT and is constructed on a 12-point interval scale, ranging from 0.00 (F) to 4.33 (A+).

College Grades. First-year GPA and grades in all courses in the first year of college were obtained from the participating institutions. All courses were coded for content area so that analyses could be conducted on course-specific grade point averages. Course-specific grade point averages were calculated within student, across all relevant course grades received in a particular area during the first semester of college (excluding remedial course work). For example, if a student took only one mathematics course in his or her first semester, then his or her average course grade in mathematics is based on the grade earned in that one course. If he or she took three mathematics courses, the average course grade is based on the average of the three course grades earned (taking into account the grades earned and the number of credits associated with each grade).

Analysis

The focus of the current study is on providing validity evidence for the use of redesigned SAT scores for college admission. Therefore, analyses were primarily correlational in nature and also graphical, depicting the relationships between the test scores and criteria of interest.

Correlational analyses were conducted to examine the strength of the relationship between the predictors of interest in the study (SAT scores and HSGPA) with FYGPA or college course grades. A correlation represents the extent to which two variables are linearly related and is on a scale of -1 to +1, where +1 is a perfect positive linear association and -1 is a perfect negative linear association. It is also helpful to think of a correlation as the extent to which a scatterplot of the relationship between two variables (e.g., SAT scores and FYGPA) fits a straight line (Miles & Shevlin, 2001).

Perfect linear associations essentially do not exist in applied social science research, so to contextualize the strength of correlation coefficients it is most helpful to either compare

correlation coefficients to other correlations representing familiar or similar relationships (Meyer et al., 2001) or refer to a rule of thumb offered by Cohen (1988). Cohen's heuristic provides a quick way to evaluate the meaningfulness of an association or effect. Correlations that have an absolute value of approximately 0.1 are considered "small," correlations that have an absolute value of approximately 0.3 are considered "medium," and correlations that have an absolute value of 0.5 or greater are considered "large." Note that correlation coefficients (corrected for restriction of range) representing relationships between admission test scores and performance in college or graduate school tend to be in the .40s and .50s (Kuncel & Hezlett, 2007; Mattern & Patterson, 2014).

Bivariate and multiple correlations in this study were calculated, and then these resulting correlation coefficients were also corrected for range restriction (both raw and corrected correlations are reported in this study). Admission validity research typically employs a correction for restriction of range because the variability of a set of predictors (e.g., SAT scores and HSGPA) is reduced due to direct or indirect selection on all or a subset of predictors. By definition, the narrowing of a score range by selection results in an underestimation of the true relationship between the predictor(s) and criterion (e.g., FYGPA). Mattern, Kobrin, Patterson, Shaw, and Camara (2009) noted that because applicants with higher HSGPAs or SAT scores are more likely to be admitted, the range of HSGPAs and SAT scores is restricted compared to the range for the full applicant pool with those measures available. This study used the Pearson-Lawley multivariate correction for restriction of range (Gulliksen, 1950; Lawley, 1943; Pearson, 1902) with the 2014 College-Bound Seniors cohort to develop the unrestricted population covariance matrix for the correction.

Separate restriction-of-range-corrected bivariate correlation matrices were computed for each participating institution instead of across all participating institutions. These separate matrices were then used to calculate the multiple correlations between the predictors and criterion as well as the average bivariate and multiple correlations, which were weighted by institution sample size.

Of particular interest in this study were the relationships between the different SAT scores (as well as all SAT section scores together) and FYGPA, as well as the incremental or additional validity that is added by the SAT to the HSGPA–FYGPA relationship. This latter relationship is estimated by examining the difference between the HSGPA–FYGPA correlation and the multiple correlation of SAT and HSGPA together with FYGPA. When possible and appropriate, relationships between SAT scores and criteria of interest are also presented graphically to more clearly show trends and relationships.

Results

First-Year Grade Point Average

Descriptive statistics for the academic variables were calculated for the student sample. Table 3 shows that this is an academically strong sample with a mean HSGPA of 3.85 and mean SAT section scores of 621 (*SD* = 100) for EBRW and 634 for Math (*SD* = 113). The mean FYGPA for the study sample was 3.30 (*SD* = 0.60). For reference, in the 2012 SAT Validity Study sample, the mean HSGPA was 3.62 (*SD* = 0.50) and the mean FYGPA was 3.02 (*SD* = 0.72).

Table 3.

Descriptive Statistics for Study Variables

	M	*SD*	*Min*	*Max*
HSGPA	3.85	0.43	1.67	4.33
SAT Total Score	1254	201	570	1600
SAT Evidenced-Based Reading and Writing Section	621	100	290	800
Reading Test	31	5.3	15	40
Writing and Language Test	31	5.2	11	40
SAT Math Section	634	113	230	800
Math Test	32	5.7	11.5	40
SAT Analysis in Science	31	5.2	13	40
SAT Analysis in History/Social Studies	31	5.3	12	40
FYGPA	3.30	0.60	0.00	4.17

Note. n = 2,050.

Table 4 shows the intercorrelation matrix for the primary predictors of interest for this study. HSGPA is correlated with both SAT sections (.50 for EBRW and .49 for Math), indicating that there is a strong relationship between the two measures but that they are not precisely measuring the same thing.

Table 4.

Corrected (Raw) Correlation Matrix of Redesigned SAT Sections and HSGPA

	HSGPA	SAT EBRW	SAT Math
HSGPA			
SAT EBRW	0.50 (0.23)		
SAT Math	0.49 (0.23)	0.77 (0.60)	

Note. n = 2,050. Restriction-of-range-corrected correlations are presented. The raw correlations are shown in parentheses.

Table 5 depicts the corrected and raw correlations of the study predictors with the primary outcome of interest in this study, the FYGPA. Confidence intervals for the corrected correlations are also presented to display the range of correlations within which we would expect the population correlation to be found with 95% confidence. Based on Cohen's (1988) rules of thumb for interpreting correlation coefficients presented earlier, you can see that the correlations between HSGPA and SAT scores with FYGPA are large, with the strongest relationship represented by the multiple correlation of both HSGPA and SAT together (r = 0.58). In this sample, the multiple correlation of the SAT EBRW and Math sections together with FYGPA is 0.53, while the correlation between HSGPA alone and FYGPA is 0.48.

To more easily understand what a correlation of 0.53 represents, you can examine Figure 1, which shows the average FYGPA that students earn by SAT total score band. In this figure, it is clear that as the SAT score band increases, there are corresponding increases in mean FYGPA. For example, those students with an SAT score between 800 and 990 earned, on average, a FYGPA of 2.89, while those students with an SAT score between 1400 and 1600 earned, on average, a FYGPA of 3.58.

Table 5.

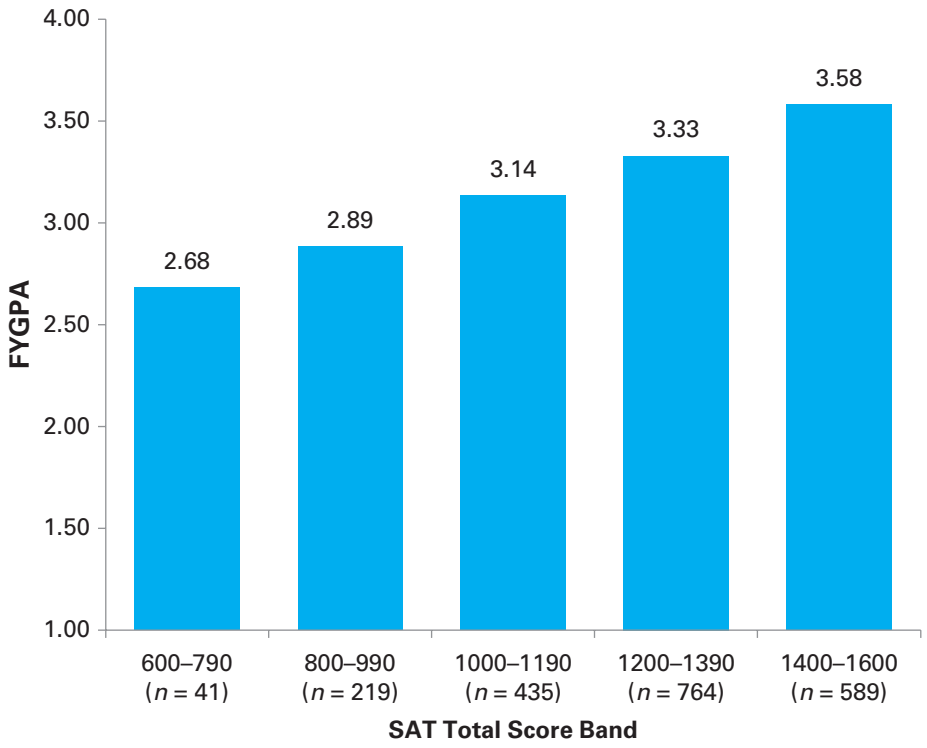
Correlations of Predictors with FYGPA

Predictors	Correlations	95% Confidence Interval for Corrected Correlations
HSGPA	0.48 (0.27)	[0.45, 0.51]
SAT EBRW Section Score	0.51 (0.33)	[0.48, 0.54]
SAT Math Section Score	0.49 (0.30)	[0.46, 0.52]
SAT EBRW, SAT Math	0.53 (0.35)	[0.50, 0.56]
HSGPA, SAT EBRW, SAT Math	0.58 (0.40)	[0.55, 0.60]

Note. n = 2,050. Pooled within institution, restriction-of-range-corrected correlations are presented. The raw correlations are shown in parentheses. The confidence intervals for bivariate correlations were calculated using the Fisher's Z transformation. Confidence intervals for the multiple correlations were calculated using the MBESS package in R.

Figure 1.

Mean FYGPA by SAT total score band.



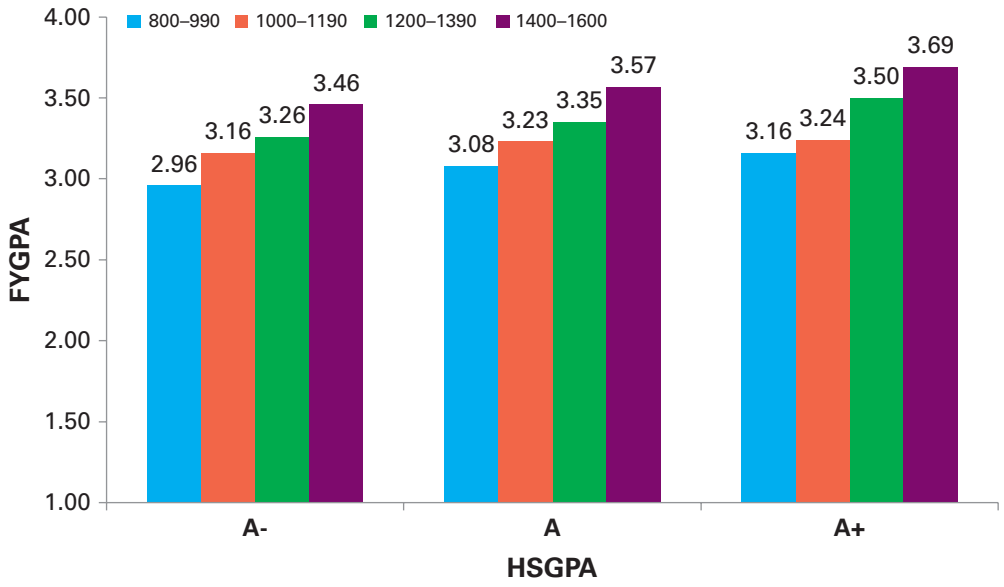
Note. Results based on fewer than 15 students are not reported (e.g., score band 400–590, n = 2).

Note that the incremental validity added by the SAT above HSGPA is 0.10 (calculated from the difference between the multiple correlation of SAT and HSGPA with FYGPA of 0.58 and the HSGPA correlation with FYGPA of 0.48). To more easily understand what this incremental validity of 0.10 represents, Figure 2 graphically depicts the mean FYGPA by SAT total score band, after controlling for HSGPA by grouping students into the same HSGPA categories (among all students who received an A). In this figure, you can see that

even within students grouped by the same HSGPAs of A-, A, or A+ (representing 84% of the study sample), there is a clear positive relationship between the SAT score bands and mean FYGPA. If there were no added value to having the SAT in addition to HSGPA in understanding students' FYGPAs, you would expect that all SAT score bands within HSGPA would have the same mean FYGPA value. Instead, for example, you can see that among those students with an "A" HSGPA, those in the SAT total score band of 800–990 have a mean FYGPA of 3.08, while those same "A" students in the SAT total score band of 1400–1600 have a mean FYGPA of 3.57.

Figure 2.

Mean FYGPA by SAT total score band, controlling for HSGPA.



Note. HSGPA ranges are defined as follows: A- = 3.67 (or 90–92), A = 4.00 (or 93–96), and A+ = 4.33 (or 97–100). Results based on fewer than 15 students are not reported; not reported are all students in the 400–590 and 600–790 SAT total score bands.

While the pre-2016 SAT was designed to maximize prediction rather than to most accurately cover a content or skills domain relevant to college, the redesigned SAT was designed to cover the content/skills that research tells us matter most to college readiness — first and foremost. In redesigning the SAT in this manner, this pilot validity study now shows us that in addition to accomplishing the desired research-based content and skills coverage, the redesigned SAT is as predictive of college success as the previous SAT. Table 6 shows the comparisons between correlation coefficients for the redesigned SAT and the pre-2016 SAT, as well as HSGPA, with FYGPA. Based on the information in Table 6, as well as the confidence intervals presented for the correlations in Table 5, we can see that the redesigned SAT correlations with FYGPA maintain the strong predictive validity of the pre-2016 SAT scores with FYGPA. This is true for the section scores as well as for the multiple correlation of SAT EBRW and Math with FYGPA. There is, however, a difference in the HSGPA–FYGPA correlation in this study sample (0.48) and in our previous validity research (0.53). Future research will need to examine whether this is a stable finding of decreased validity for the HSGPA or if it is sample specific. However, this lower HSGPA correlation with FYGPA also

Table 6.

Comparison of Correlations of Predictors with FYGPA from Redesigned SAT Pilot Validity Study Sample and 2012 National SAT Validity Study Sample

Redesigned SAT Pilot Validity Study *n* = 2,050		National SAT (pre-2016) Validity Study: 2012 cohort *n* = 223,109	
Predictors	**Correlations**	**Correlations**	**Predictors**
HSGPA	0.48 (0.27)	0 .53 (0.34)	HSGPA
SAT EBRW Section Score	0.51 (0.33)	0.48 (0.27)	SAT Critical Reading Section Score
		0.52 (0.33)	SAT Writing Section Score
SAT Math Section Score	0.49 (0.30)	0.48 (0.26)	SAT Math Section Score
SAT EBRW, SAT Math	0.53 (0.35)	0.54 (0.35)	SAT Critical Reading, Writing, Math
HSGPA, SAT EBRW, SAT Math	0.58 (0.40)	0.61 (0.44)	HSGPA, SAT Critical Reading, Writing, Math

Note. Pooled within institution, restriction-of-range-corrected correlations are presented. The raw correlations are shown in parentheses.

impacts the multiple correlation of HSGPA and SAT with FYGPA, whereby the lower HSGPA–FYGPA correlation brings down multiple correlation in this study (0.58) as compared to previous research (0.61).

Course-Specific Grade Point Average

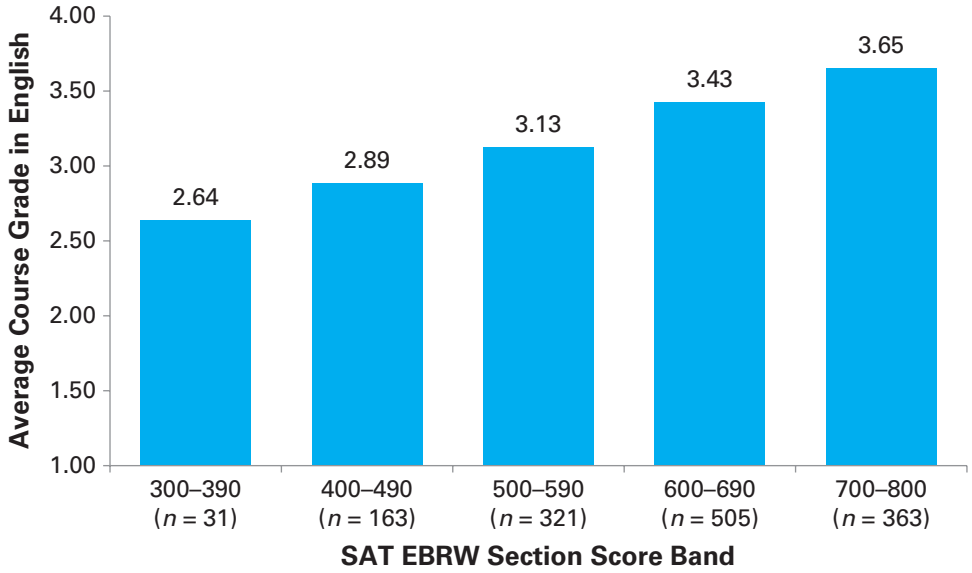
In addition to understanding the relationships between SAT scores and FYGPA based on correlational analysis, we explored the relationships between SAT section and cross-test scores with average first-semester course grades in the matching domain using graphical representations. All student course work data in this study were coded for their content area focus as well as for whether or not they were from remedial courses. Remedial course work was not included in this analysis.

Content experts, assessment developers, and researchers then worked to match the appropriate course work codes with the matching SAT scores so that the relationship between the scores and college performance in the matching content area could be examined. Figure 3 shows the relationship between SAT EBRW scores and average first-semester credit-bearing college course grades in reading- and writing-intensive courses, including history, literature (not composition), social science, and writing courses. This graph depicts a clear positive relationship between SAT EBRW scores and grades in matching college courses. For example, those students with an SAT EBRW score of 400–490 have an average matching college course grade of 2.89, whereas those students with an SAT EBRW score of 700–800 have an average matching college course grade of 3.65.

Figure 4 shows the relationship between SAT Math scores and average first-semester credit-bearing college course grades in algebra, precalculus, calculus, and statistics. This graph depicts a clear positive relationship between SAT Math scores and grades in matching college courses. For example, those students with an SAT Math score of 400–490 have an average matching college course grade of 2.50, whereas those students with an SAT Math score of 700–800 have an average matching college course grade of 3.27.

Figure 3.

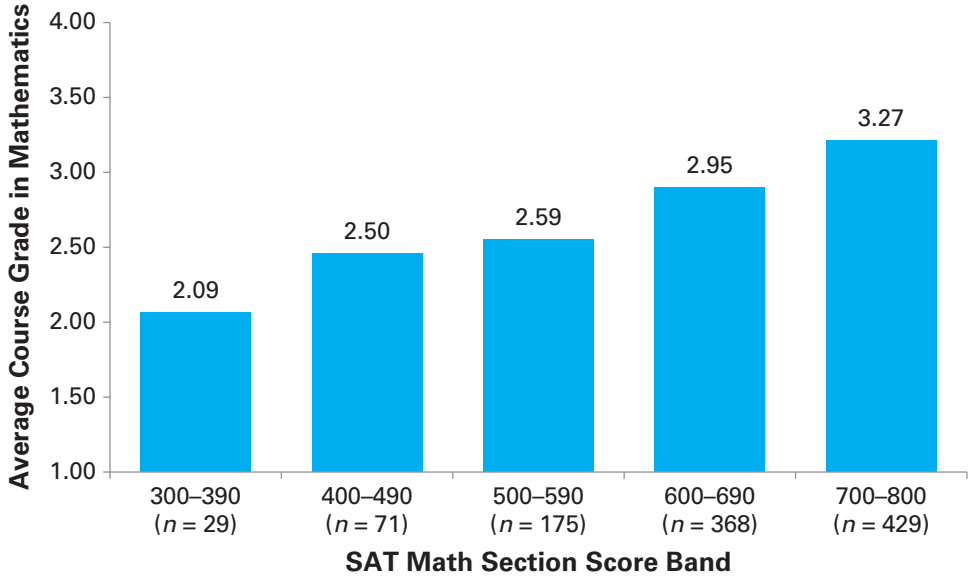
Relationship between SAT Evidence-Based Reading and Writing scores and course grades in the same domain.



Note. Results based on fewer than 15 students are not reported (e.g., score band 200–290, $n = 1$). Average English course grade includes first-semester courses that are reading and writing intensive (excluding foreign and classical languages).

Figure 4.

Relationship between SAT Math Section scores and course grades in the same domain.

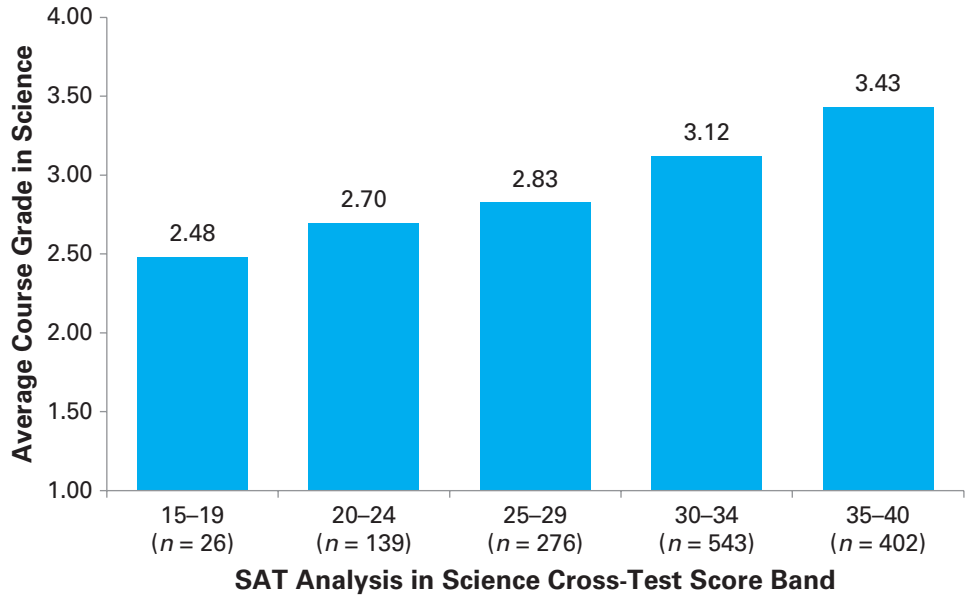


Note. Results based on fewer than 15 students are not reported (e.g., score band 200–290, $n = 1$). Average math course grade includes first-semester course work in algebra, precalculus, calculus, and statistics.

Figure 5 shows the relationship between SAT Analysis in Science cross-test scores and average first-semester credit-bearing college course grades in science, including natural sciences, health sciences, and engineering. This graph depicts a clear positive relationship between SAT Analysis in Science cross-test scores and grades in matching college courses. For example, those students with an SAT Analysis in Science cross-test score of 20–24 have an average matching college course grade of 2.70, whereas those students with an SAT Analysis in Science cross-test score of 35–40 have an average matching college course grade of 3.43.

Figure 5.

Relationship between SAT Analysis in Science cross test scores and course grades in the same domain.

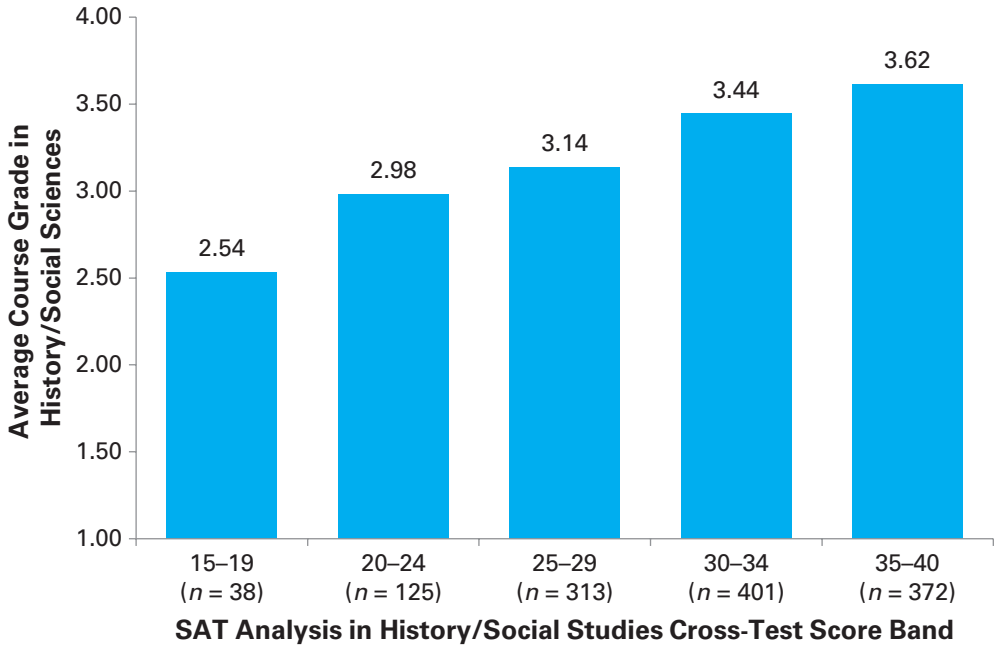


Note. Results based on fewer than 15 students are not reported (e.g., score band 10–14, $n = 0$). Average science course grade includes first-semester course work in natural sciences, health sciences, and engineering.

Figure 6 shows the relationship between SAT Analysis in History/Social Studies cross-test scores and average first-semester credit-bearing college course grades in history (e.g., world history, U.S. history, European history, etc.) and social sciences (e.g., anthropology, economics, government, geography, psychology, etc.) course work. This graph depicts a clear positive relationship between SAT Analysis in History/Social Studies cross-test scores and grades in matching college courses. For example, those students with an SAT Analysis in History/Social Studies cross-test score of 20–24 have an average matching college course grade of 2.98, whereas those students with an SAT Analysis in History/Social Studies cross-test score of 35–40 have an average matching college course grade of 3.62.

Figure 6.

Relationship between SAT Analysis in History/Social Studies cross-test scores and course grades in the same domain.



Note. Results based on fewer than 15 students are not reported (e.g., score band 10–14, $n = 3$). Average history/social studies course grade includes first-semester course work in history and social sciences.

Discussion

This pilot predictive validity study of the redesigned SAT allowed for a preliminary look at the relationship between redesigned SAT scores and grades in the first year of college. Across a diverse sample of first-year students at 15 four-year institutions, the results of this pilot study showed that redesigned SAT scores remain as predictive of college success as pre-2016 SAT scores. This is important to note as the redesign of the SAT was first and foremost focused on more closely aligning the content and skills tested on the SAT with those that research indicates are critical for college success. In making these important changes to the test, the fact that the strong predictive validity was also maintained is a significant accomplishment of the redesign.

In addition, this study showed that redesigned SAT scores improve the ability to predict college performance above high school GPA alone — and more so than in previous studies. In other words, while the SAT and HSGPA are both measures of a student's previous academic performance that are strongly related to FYGPA in college, they also tend to measure somewhat different aspects of academic performance and therefore complement each other in their use in college admission and the overall prediction of FYGPA.

Finally, the examination of the relationships between the SAT section scores and cross-test scores with grades in the matching course work domain(s) in college shows a strong positive

relationship, suggesting that the redesigned SAT is sensitive to instruction in English/language arts, math, science, and history/social studies. Just as one would expect, higher SAT section or cross-test scores are associated with higher course grades in the matching academic field in college.

As with all pilot studies that include a pilot form of the test, a smaller sample, and students who may be less motivated to perform at their best, it will be important to replicate the study findings with a large, nationally representative sample after the redesigned SAT becomes operational. The College Board will be launching such a study, examining students in the entering college class of fall 2017, or the first full cohort to be admitted to college with the redesigned SAT. These students will complete one year of college, and then in the fall of 2018 and the year that follows we will be able to study the relationship between redesigned SAT scores and first-year college performance. We will continue to track students through college so that relationships between redesigned SAT scores and longer-term outcomes such as persistence, completion, and cumulative GPA can be studied.

While validity research is critical to conduct and disseminate to test users at the national level, it is also important for institutions to continue to conduct their own local validity research studies examining the relationship between SAT scores with college grades. The College Board offers a free online validity study service to help institutions with this endeavor called the Admitted Class Evaluation Service™ (ACES™). ACES provides institutions with their own validity studies, uniquely tailored to their institution; however, institutions can always conduct such studies on their own and may be able to reference the design utilized and decisions made in this current study to assist in their own work in this area.

References

Beard, J., & Marini, J. P. (2015). *Validity of the SAT for predicting first-year grades: 2012 SAT validity sample* (College Board Statistical Report No. 2015-2). New York: The College Board.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

College Board. (2014). *2014 College-bound seniors: Total group profile report*. New York: The College Board. Retrieved from: https://secure-media.collegeboard.org/digitalServices/pdf/sat/TotalGroup-2014.pdf

Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley and Sons.

Kelly, K., & Lai, K. (2012). MBESS: MBESS. R package version 3.3.3. [Computer Software]. http://CRAN.R-project.org/package=MBESS

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science, 315*(5815), 1080–1081.

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh*, Section A, *62*(1), 28–30.

Mattern, K. D., Kobrin, J. L., Patterson, B. F., Shaw, E. J., & Camara, W. J. (2009). Validity is in the eye of the beholder: Conveying SAT research findings to the general public. In R.W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 213–240). Charlotte, NC: Information Age Publishing.

Mattern, K. D., & Patterson, B. F. (2014). *Synthesis of recent SAT validity findings: Trend data over time and cohorts* (College Board Research in Review 2014-1). New York: The College Board. Retrieved from: http://research.collegeboard.org/sites/default/files/publications/2014/6/Synthesis-of-Recent-SAT-Validity-Findings.pdf

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J., Kubiszyn, T. W., & Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, *56*(2), 128–165.

Miles, J., & Shevlin, M. (2001). *Applying regression & correlation: A guide for students and researchers*. London: Sage Publications.

Pearson, K. (1902). *Philosophical Transactions of the Royal Society of London, Series A, Containing Papers of a Mathematical or Physical Character*, pp. 198, 235–299.

R Development Core Team (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org

The Research department actively supports the College Board's mission by:

- Providing data-based solutions to important educational problems and questions
- Applying scientific procedures and research to inform our work
- Designing and evaluating improvements to current assessments and developing new assessments as well as educational tools to ensure the highest technical standards
- Analyzing and resolving critical issues for all programs, including AP®, SAT®, PSAT/NMSQT®
- Publishing findings and presenting our work at key scientific and education conferences
- Generating new knowledge and forward-thinking ideas with a highly trained and credentialed staff

Our work focuses on the following areas

Admission	Measurement
Alignment	Research
Evaluation	Trends
Fairness	Validity

Follow us online: research.collegeboard.org

CollegeBoard

CollegeBoard

00078 011